

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
1 avril 2004 (01.04.2004)

PCT

(10) Numéro de publication internationale
WO 2004/027066 A2

(51) Classification internationale des brevets⁷ : C12N 15/62

(21) Numéro de la demande internationale :
PCT/FR2003/002712

(22) Date de dépôt international :
15 septembre 2003 (15.09.2003)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
02/11485 17 septembre 2002 (17.09.2002) FR

(81) États désignés (*national*) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) États désignés (*régional*) : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Déposant (*pour tous les États désignés sauf US*) :
BIOMERIEUX [FR/FR]; Département Propriété Industrielle, Chemin de l'Orme, F-69280 Marcy l'Etoile (FR).

Déclaration en vertu de la règle 4.17 :

— *relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement*

(72) Inventeur; et

Publiée :

(75) Inventeur/Déposant (*pour US seulement*) : LE-
TOURNEUR, Odile [FR/FR]; 6 Allée Buffon, F-69110
SAINTÉ FOY LES LYON (FR).

— *sans rapport de recherche internationale, sera republiée dès réception de ce rapport*

(74) Mandataire : BIOMERIEUX; Mademoiselle DENJEAN
Frédérique, Département Propriété Industrielle, Chemin de
l'Orme, F-69280 Marcy l'Etoile (FR).

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(54) Title: CHIMERIC RECOMBINANT PROTEIN AND IN VITRO DIAGNOSIS

(54) Titre : PROTÉINE RECOMBINANTE CHIMÉRIQUE ET DIAGNOSTIC IN VITRO

(57) Abstract: The invention relates to recombinant DNA coding for a chimeric recombinant protein comprising at least two first nucleotide fragments, each coding an epitopic region of the HIV -1 virus of group M or group O or the virus HIV-2, at least one second nucleotide fragment coding a linking region, at least one third nucleotide fragment coding a fixing region, characterized in that each first nucleotide fragment codes at least one immunodominant region of glycoprotein gp 120 of HIV-1, of glycoprotein gp 41 of HIV-1 of group M, of glycoprotein gp 41 of HIV-1 of group O or glycoprotein gp 36 of HIV-2. The invention also contains a recombinant chimeric protein coded by the above-mentioned DNA, and the use of said DNA and/or said recombinant protein for in vitro diagnosis.

(57) Abrégé : L'invention concerne un ADN recombinant codant une protéine recombinante chimérique comprenant - au moins deux premiers fragments nucléotidiques codant chacun une région épitopique du virus HIV- 1 de groupe M ou de groupe O ou le virus HIV-2 - au moins un deuxième fragment nucléotidique codant une région de liaison, - au moins un troisième fragment nucléotidique codant une région de fixation caractérisé en ce que chaque premier fragment nucléotidique code au moins une région immunodominante de la glycoprotéine gp 120 d'HIV-1, de la glycoprotéine gp 41 d'HIV-1 de groupe M, de la glycoprotéine gp 41 d'HIV-1 de groupe O ou de la glycoprotéine gp 36 d'HIV-2. L'invention concerne également une protéine chimérique recombinante codée par l'ADN défini ci dessus, ainsi que l'utilisation dudit ADN et/ou de ladite protéine recombinante pour le diagnostic in vitro.

Protéine recombinante chimérique et diagnostic in vitro

La présente invention est relative à une protéine recombinante chimérique, un ADN codant ladite protéine recombinante chimérique, ainsi que l'utilisation de cette protéine recombinante chimérique pour le diagnostic in vitro des maladies en relation avec un virus, plus particulièrement le virus HIV-1 et/ou et HIV-2.

Le diagnostic précoce de la présence d'un virus dans l'organisme est essentiel afin, d'une part, d'éviter la propagation du virus par les malades qui ne connaissent pas encore leur séropositivité, et d'autre part, de proposer à ces patients un traitement adapté pour reculer la date d'apparition des symptômes.

Dans le cas du SIDA (syndrome d'immunodéficience acquise), conséquence de l'infection par les rétrovirus HIV-1 (human immunodeficiency virus-1) ou HIV-2 (human immunodeficiency virus-2), la primo infection est suivie d'une période asymptomatique, d'une durée variable avant que la maladie n'évolue chez la plupart des patients en SIDA, caractérisée par l'apparition d'infections à germes opportunistes, de tumeurs et de manifestations neurologiques, et le diagnostic précoce de la présence du virus HIV dans l'organisme doit être effectué que le patient ait initialement été infecté par HIV-1 ou par HIV-2.

La plupart des tests de diagnostic commercialisés est basé sur une réaction antigène-anticorps dirigée contre certaines protéines virales, telles que les protéines transmembranaires de l'enveloppe virale. Dans le cas d'HIV-1, les protéines de l'enveloppe sont dérivées du gène env, qui code une glycoprotéine précurseur d'un poids moléculaire de 160 000 daltons, dénommée gp 160. La gp 160 est ensuite clivée en deux protéines virales de l'enveloppe, la gp 120 et la gp 41. Dans le cas d'HIV-2, la glycoprotéine précurseur est la gp140, clivée en gp36 et gp105/110. Ainsi, dans l'article scientifique de Vallari *et al* (Journal of microbiology, pages 3657-3661, 1998), on retrouve la description d'un kit de diagnostic permettant de détecter la présence de virus HIV-1 groupe M, HIV-1 groupe O et HIV-2, par l'utilisation de trois protéines recombinantes, dérivées des régions env de virus HIV-1 groupe M, HIV-1 groupe O et HIV-2. D'une façon comparable, dans l'article scientifique de Shin *et al* (Biochemistry

and molecular biology international, volume 43, n°4, pages 713-721, 1997), sont décrits des peptides multi-antigéniques pour détecter des infections aux virus HIV-1 ou HIV-2.

Toutefois, dans de tels tests de diagnostic de plusieurs souches virales, il est nécessaire lors de l'analyse de séra par la technique ELISA (Enzyme Linked ImmunoSorbent Assay), de fixer sur le support plusieurs protéines recombinantes ayant des caractéristiques d'adsorption (ou coating) individuelles différentes, induisant des problèmes notamment lors de l'étape de révélation. De plus, la multiplicité des protéines recombinantes engendre des coûts de fabrication importants.

Afin d'éviter ces problèmes de différences de coating entre les différentes protéines recombinantes, d'autres tests de diagnostic utilisent préférentiellement des protéines recombinantes chimériques, porteuses de plusieurs épitopes dirigés contre des protéines virales différentes.

Ainsi, le brevet EP-B-0 577 894 décrit la construction d'une protéine recombinante chimérique utilisée pour le diagnostic du SIDA. Cette protéine est porteuse des épitopes dirigés contre les protéines virales dérivées du gène gag de HIV-2 et contre la protéine gp120 d'HIV-1. Toutefois, cette protéine recombinante ne permet pas la détection simultanée des patients infectés par les virus HIV-1 de groupe M et O, ce qui peut induire des risques de faux négatifs (patient détecté séronégatif alors qu'il est porteur du virus), faux négatifs dont les conséquences peuvent être dramatiques. De plus, cette protéine recombinante n'est pas porteuse de l'épitope dirigé contre la gp41, qui est pourtant l'épitope immunodominant majeur, ce qui augmente, là encore, le risque d'apparition de faux négatif. De même, dans l'article scientifique de Han *et al* (Biochemistry and molecular biology international, vol 46 n°3, 1998), il est décrit une protéine recombinante présentant un épitope dirigé contre la gp41 d'HIV-1, et un épitope dirigé contre la gp36 d'HIV-2, ces deux épitopes étant liés par un peptide de liaison afin de permettre l'accessibilité à chacun des épitopes. Toutefois, cette protéine recombinante ne permet pas la détection simultanée des patients infectés par les virus HIV-1 de groupe M et O, ce qui peut induire, là encore, des risques de faux négatifs. La demande de brevet DE 101 06 295 décrit une protéine recombinante comprenant plusieurs épitopes dirigés contre HIV-1 ou HIV-2, liés par des régions de liaisons, permettant d'immobiliser la protéine recombinante sur un support solide. Toutefois, les

régions épitopiques de cette protéine recombinante permettent la reconnaissance d'anticorps dirigés contre les produits du gène pol du virus HIV (protéase, reverse transcriptase ou endonucléase) ou contre la séquence qui constitue la boucle V3 de la gp120. Les protéines codées par le gène pol étant relativement conservées d'un virus à un autre, les anticorps dirigés contre ces protéines sont peu spécifiques. De plus, les anticorps dirigés contre les antigènes pol d'un virus apparaissent chez les personnes infectées tardivement ce qui ne permet pas un diagnostic de la maladie en début d'infection. Il a été montré également que le titre en anticorps anti-antigènes pol diminue avec la progression de la maladie, et qu'en conséquence, un diagnostic faussement négatif pouvait être attribué à un malade en phase d'infection chronique. En ce qui concerne la séquence de la boucle V3 de la gp120, cette séquence est hypervariable et l'utilisation de 2 ou 3 séquences dites sous types spécifiques ne garantit pas la détection de tous les anticorps dirigés contre ce domaine : des résultats faussement négatifs peuvent donc être obtenus.

15

La présente invention se propose de résoudre l'ensemble des inconvénients de l'état de la technique, en proposant une nouvelle protéine recombinante chimérique, facile à purifier et à synthétiser, présentant une forte immunoréactivité vis-à-vis de séra de patients susceptibles d'être infectés par un ou plusieurs virus, tel que HIV-1, groupe M et ou O ou HIV-2.

20

Les définitions suivantes permettront de mieux comprendre l'invention.

Par ADN recombinant, on entend une séquence nucléotidique construite artificiellement et obtenue par génie génétique. A titre indicatif, ledit ADN recombinant peut être inséré dans un organisme hôte d'expression, tel que notamment une bactérie, par un vecteur d'expression, notamment un plasmide bactérien ou un bactériophage.

25

Par fragment nucléotidique, on entend une succession d'au moins trois acides nucléotidiques codant au moins un acide aminé.

Par site de coupure, on entend un site permettant la séparation de deux fragments nucléotidiques par l'action d'au moins un moyen de coupure, tel que notamment une enzyme de restriction qui est capable, au niveau de site de coupure correspondant à une

30

séquence nucléotidique spécifique de générer pour chaque brin deux extrémités, l'une ayant un groupement 3'-OH, l'autre un groupement 5'-P.

Par protéine recombinante chimérique, on entend une protéine construite artificiellement et obtenue par génie génétique. A titre indicatif, ladite protéine recombinante chimérique peut être produite par un organisme hôte d'expression modifiée génétiquement par insertion de la séquence nucléotidique codant ladite protéine recombinante chimérique par un vecteur d'expression.

Par région épitopique, on entend une région peptidique qui va interagir de façon stéréospécifique avec la région peptidique paratopique de l'anticorps dirigé contre le microorganisme, tel que notamment le virus, présent dans le sérum du patient. Les régions épitopiques les plus immunogéniques sont dites immunodominantes.

Par région de liaison, on entend une région assurant une meilleure accessibilité des régions paratopiques des différents anticorps présents dans le sérum des patients vis à vis des régions épitopiques d'intérêt correspondantes de la protéine recombinante chimérique.

Par région de fixation, on entend une région permettant de fixer ladite protéine recombinante chimérique vis-à-vis :

- d'un support, de manière direct et/ou indirect, et/ou
- d'une molécule de détection,

Le support peut être composé de matériaux, tels que :

- le verre, un matériau peu onéreux, inerte et mécaniquement stable ,
- les polymères : plaques de microtitration...
- les métaux : colonne de métal chelate de chromatographie d'affinité,
- les particules magnétiques, telles que décrites dans les demandes de brevet WO-A-97/34909, WO-A-97/45202, WO-A-98/47000 et WO-A-99/35500 déposées par la demanderesse.

Le support peut alors être utilisé comme support d'analyse, notamment lors d'un ELISA (Enzym Linked ImmunoSorbent Assay), pour des étapes de purification lors d'une chromatographie d'affinité, pour des étapes de lavage lorsque ladite protéine recombinante chimérique, fixée sur une particule magnétique est retenue par aimantation dans un lieu prédéterminé.

Par molécule de détection, on entend une molécule associée à un marqueur permettant de générer directement ou indirectement un signal détectable. Ces marqueurs peuvent être notamment radioactifs, enzymatiques, fluorescents.

La fixation de la protéine recombinante chimérique sur ledit support ou ladite molécule de détection peut impliquer des ligands capables de réagir avec un anti-ligand. On peut citer à titre d'exemples les couples ligand/anti-ligand suivants :

- biotine/streptavidine,
- haptène/anticorps,
- antigène/anticorps,
- 10 • peptide/anticorps,
- sucre/lectine,

Ainsi, l'invention concerne un ADN recombinant codant une protéine recombinante chimérique comprenant

- 15 ☐ au moins deux premiers fragments nucléotidiques codant chacun une région épitopique du virus HIV-1 de groupe M ou de groupe O ou du virus HIV-2
 - ☐ au moins un deuxième fragment nucléotidique codant une région de liaison,
 - ☐ au moins un troisième fragment nucléotidique codant une région de fixation
- caractérisé en ce que chaque premier fragment nucléotidique code au moins une région immunodominante de la glycoprotéine gp 120 d'HIV-1, de la glycoprotéine gp 41 d'HIV-1 de groupe M, de la glycoprotéine gp 41 d'HIV-1 de groupe O ou de la glycoprotéine gp 36 d'HIV-2.

Il est bien évidemment que les régions variables de ces région immunodominante telle que notamment la boucle V3 de la gp120 ne sont nullement envisagées dans la présente invention.

Selon un mode préféré de réalisation de l'invention, ledit premier fragment nucléotidique a pour séquence l'une quelconque des séquences SEQ ID N°3, SEQ ID N°5, SEQ ID N°7, SEQ ID N°9, SEQ ID N°27, SEQ ID N°29 ou SEQ ID N°31.

Selon un autre mode préféré de réalisation de l'invention, ledit deuxième fragment nucléotidique comprend au moins un site de coupure. Préférentiellement, ledit deuxième fragment nucléotidique a pour séquence au moins l'une quelconque des

séquences suivantes, prises seules ou en combinaison, SEQ ID N°11 SEQ ID N°13, SEQ ID N°15, SEQ ID N°17, SEQ ID N°19, ou SEQ ID N°20, SEQ ID N°33, SEQ ID N°35, SEQ ID N°37, SEQ ID N°39, SEQ ID N°41, SEQ ID N°43 SEQ ID N°45, ou SEQ ID N°47

5 Selon un mode préféré de réalisation de l'invention, ledit troisième fragment nucléotidique codant une région de fixation est inclus dans ledit deuxième fragment nucléotidique codant une région de liaison.

Selon un autre mode préféré de réalisation de l'invention, ledit troisième fragment nucléotidique a pour séquence l'une quelconque des séquences SEQ ID N°21, SEQ ID
10 N° 23 23, ou SEQ ID N°25, SEQ ID N°33, SEQ ID N°35, SEQ ID N°37 ou SEQ ID N°39.

Les séquences nucléotidiques selon l'invention peuvent être préparées par synthèse chimique et génie génétique en utilisant les techniques bien connues de l'homme du métier et décrites par exemple dans Sambrook J. et al., Molecular
15 Cloning : A Laboratory Manual, 1989.

Les séquences nucléotidiques de l'invention peuvent être insérées dans des vecteurs d'expression afin de préparer les protéines recombinantes de l'invention.

L'invention concerne également une protéine recombinante chimérique codée par un
20 ADN recombinant, tel que défini précédemment, comprenant

- ☐ au moins deux régions épitopiques du virus HIV-1 de groupe M ou de groupe O ou le virus HIV-2 d'au moins un microorganisme,
- ☐ au moins une région de liaison
- ☐ au moins une région de fixation.

25 Selon un mode préféré de réalisation de l'invention, ladite région de liaison est un peptide comprenant au moins une glycine et/ou au moins une sérine.

Selon un autre mode préféré de réalisation de l'invention, ladite région de liaison a pour séquence l'une quelconque des séquences SEQ ID N° 12, SEQ ID N°14, SEQ ID N°16 ou SEQ ID N°18, 34, 36, 38, 40, 42, 44, 46, 48.

30 Selon un autre mode de réalisation de l'invention, ladite région de fixation est une région riche en histidines et ses dérivés, tel qu'une région contenant une densité

d'histidines, supérieure ou égale à 25%, et préférentiellement supérieure ou égale à 33%.

Selon un autre mode de réalisation de l'invention, ladite région de fixation est un peptide comprenant au moins une lysine.

- 5 Selon un mode préféré de réalisation de l'invention, ladite région de fixation a pour séquence SEQ ID N°22, 24 ou , 26, 34, 36, 38 ou 40.

Les protéines recombinantes de l'invention peuvent être obtenues par la technique du génie génétique qui comprend les étapes de :

- 10 - culture d'organismes hôtes ou de cellules eucaryotes transformées à l'aide d'une séquence nucléotidique selon l'invention et
- récupération de la protéine produit par lesdits organismes hôtes ou lesdites cellules eucaryotes transformées.

Cette technique est bien connue de l'homme du métier. Pour plus de détails la concernant, on pourra se référer à l'ouvrage ci-après : Recombinant DNA Technology I,
15 Editors Ales Prokop, Raskesh K Bajpai ; Annals of the New-York Academy of Sciences, Volume 646, 1991.

L'invention concerne également un vecteur d'expression comprenant un ADN recombinant tel que défini précédemment.

- 20 A titre de vecteur d'expression, on peut citer par exemple les plasmides, les vecteurs viraux type virus de la vaccine, adenovirus, baculovirus, les vecteurs bactériens du type salmonelle, BCG.

On entend par moyen nécessaire à l'expression d'une protéine, tout moyen qui permet d'obtenir ladite protéine, tel que notamment un promoteur, un terminateur de
25 transcription, une origine de réplication et de préférence un marqueur de sélection.

Les vecteurs de l'invention peuvent également comprendre des séquences nécessaires au ciblage des protéines vers des compartiments cellulaires particuliers. Un exemple de ciblage peut être le ciblage vers le réticulum endoplasmique obtenu en utilisant des séquences d'adressage du type de la séquence leader issue de la protéine E3 de l'adénovirus (Ciernik L.F., et al., The Journal of Immunology, 1999, 162, 3915-3925).

A titre d'exemples d'organismes hôtes qui conviennent aux fins de l'invention, on peut citer les levures, telles que celles des familles suivantes : *Saccharomyces*, *Schizosaccharomyces*, *Kluveromyces*, *Pichia*, *Hanseluna*, *Yarrowia*, *Schwaniomyces*, *Zygosaccharomyces*, *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis* et *Kluveromyces lactis* étant préférées ; et les bactéries, telles que *E. coli* et celles des familles suivantes : *Lactobacillus*, *Lactococcus*, *Salmonella*, *Strptococcus*, *Bacillus* et *Streptomyces*.

A titre d'exemples de cellules eucaryotes, on peut citer les cellules provenant d'animaux tels que les mammifères, les reptiles, les insectes et équivalent. Les cellules eucaryotes préférées sont les cellules provenant du hamster chinois (cellules CHO), du singe (cellules COS et Vero), du rein de hamster nain (cellules BHK), du rein de cochon (cellules PK 15) et du rein de lapin (cellules RK13, les lignées cellulaires humaines de l'ostéosacorne (cellules 143 B), les lignées cellulaires humaines HeLa et les lignées cellulaires humaines de l'hépatome (du type cellules Hep G2), ainsi que les lignées cellulaires d'insecte (par exemple de *Spodoptera frugiperda*).

L'invention concerne enfin l'utilisation d'au moins un ADN tel que défini précédemment et/ou d'au moins une protéine recombinante chimérique telle que définie précédemment pour le diagnostic in vitro. Cette utilisation permet la détection du virus HIV-1 de groupe M et O ainsi que la détection du virus HIV-2.

Les exemples suivants sont donnés à titre illustratif et n'ont aucun caractère limitatif. Ils permettront de mieux comprendre l'invention.

Exemple 1 : Construction des protéines chimériques recombinantes b-HIV72, b-HIV86 et b-HIV98 permettant la reconnaissance d'anticorps anti HIV-1 groupe O et M et HIV-2.

La séquence nucléotidique SEQ ID N°1 a été conçue pour coder une protéine recombinante b-HIV72, et clonée dans un vecteur d'expression. Elle correspond à la séquence suivante :

SEQ ID N° 1 : ATG AGG GGA TCC AGA ATC CTA GCT GTG GAA AGA TAC
CTA AAG GAT CAA CAG CTC CTA GGG ATT TGG GGT TGC TCT GGA
AAA CTC ATT TGC ACC ACT GCT GTG AGC TCC GGT TCA GGC GCT ATA
GAG AAG TAC CTA CAG GAC CAG GCG CGG CTA AAT TCA TGG GGA
TGT GCG TTT AGA CAA GTC TGC TCG AGC GGT TCT GGA GGA GGA
GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA TAT AAA TAT AAA
GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC AAG TCT
GCA GGC CGT CTG CTT GCT CTG GAA ACC CTG CTT CAG AAC CAA
CAG CTG CTT TCT CTG TGG GGT TGC AAA GGT AAG CTG GTT TGC
TAC ACC TCT GTT AAA GCT TCC CAC CAT CAC CAT CAC CAT TGA TCT
AGA

La protéine recombinante chimérique b-HIV72 codée par la séquence SEQ ID N°1, comprend 137 acides aminés, pour une masse moléculaire de 15191,5 Da. Sa séquence en acides aminés est la suivante :

SEQ ID N°2 : MRGS RILAVERYLK DQQLLGIWGC SGKLICTTAV SSGSG
AIEKYLQDQA RLNSWGCAFR QVC SSGS GGGDMRDNWR SELYKYKVKV
IEPLGVAPTK SAG RLLALETLLQ NQQLLSLWG CKGKLV CYTS V KAS
HHHHHH.

La présence de MRGS et la séquence ATG AGG GGA TCC correspondante est introduite par la technique de clonage utilisée dans le vecteur d'expression pMR. La séquence d'intérêt est introduite dans le vecteur pMR entre le site de restriction BamHI en 5' et le site XbaI en 3', ce qui conduit à la fusion de la séquence MRGS en N-terminal de la protéine d'intérêt. Seul le codon d'initiation ATG et par conséquent l'acide aminé Met est vraiment indispensable dans cette séquence.

Les régions épitopiques sont indiquées en gras, la région de fixation en italique et les régions de liaisons en non gras non italique.

Cette protéine recombinante chimérique b-HIV72 comprend :

- 5 a) plusieurs régions épitopiques (indiquées en gras dans la SEQ ID n°2) permettant :

- la reconnaissance des anticorps anti-HIV-1 (groupe M ; gp41), :

La séquence SEQ ID N°3 est issue de la souche virale HIV-1 groupe M (clone de référence HXB2) et correspond à la séquence suivante :

SEQ ID N°3 : AGA ATC CTA GCT GTG GAA AGA TAC CTA AAG GAT CAA
10 CAG CTC CTA GGG ATT TGG GGT TGC TCT GGA AAA CTC ATT TGC ACC
ACT GCT GTG,

Cette séquence est amplifiée par PCR (polymerase chain reaction) par l'utilisation d'amorces d'amplification spécifiques (amorce sense 5' AGT CGG ATC CAG AAT CCT AGC TGT GGA A 3' et amorce antisense 5' GCC TGA TCC GGA GCT CAC
15 AGC AGT GGT GCA AAT 3' ; 17 cycles de PCR sont réalisés avec à chaque cycle une étape de dénaturation à 94°C pendant 1 minute (min), une étape d'hybridation à 52°C pendant 1 min et une étape d'élongation 72°C pendant 20 secondes.

Le fragment nucléotidique obtenu code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°4 : RILAVERYLK DQQLGIWGC SGKLICTTAV.

- 20 • la reconnaissance des anticorps anti-HIV-1 (groupe O ; gp41) :

La séquence SEQ ID N° 5 correspond à une séquence ADN artificielle, conçue à partir de la séquence en acide aminé de la souche virale HIV-1 groupe O [clone ANT70].

Cette portion synthétique a été conçue en sélectionnant des codons dont l'utilisation est favorable à l'expression de gène chez *E. coli*. La séquence est la suivante :

25 SEQ ID N° 5 : CGT CTG CTT GCT CTG GAA ACC CTG CTT CAG AAC CAA
CAG CTG CTT TCT CTG TGG GGT TGC AAA GGT AAG CTG GTT TGC TAC
ACC TCT GTT.

Cette séquence est construite par PCR par l'utilisation de 3 oligonucléotides (un oligonucléotide sens 5' AAG TCT GCA GGC CGT CTG CTT GCT CTG GAA ACC
30 CTG CTT CAG AAC CAA CAG CTG CTT TCT 3' et deux oligonucléotides antisens
5' GCT ATC TAG ATC AAT GGT GAT GGT GAT GGT GGG AAG CTT TAA CAG
AGG TGT AGC AAA C 3' et 5' AAC AGA GGT GTA GCA AAC CAG CTT ACC

TTT GCA ACC CCA CAG AGA AAG CAG CTG TTG GTT 3' ; 17 cycles de PCR sont réalisés avec à chaque cycle une étape de dénaturation à 9°C pendant 1 min, une étape d'hybridation à 50°C pendant 1 min et une étape d'élongation 68°C pendant 20 secondes).

- 5 Ce fragment nucléotidique code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°6 : RLLALETLLQ NQQLLSLWGC KGKLVCYTSV .

• la reconnaissance des anticorps anti-HIV-2 (gp36) :

La séquence SEQ ID N°7 est issue de la souche virale HIV-2 (clone de référence ROD) et correspond à la séquence suivante :

- 10 SEQ ID N°7 : GCT ATA GAG AAG TAC CTA CAG GAC CAG GCG CGG CTA
AAT TCA TGG GGA TGT GCG TTT AGA CAA GTC TGC

Cette séquence est amplifiée par PCR par l'utilisation d'amorces d'amplification spécifiques (amorce sense 5' CTG TGA GCT CCG GTT CAG GCG CTA TAG AGA AGT ACC TA 3' et amorce antisense 5' AGA ACC GCT CGA GCA GAC TTG TCT
15 AAA CGC 3' ; 17 cycles de PCR sont réalisés avec à chaque cycle une étape de dénaturation à 94°C pendant 1 min, une étape d'hybridation à 52°C pendant 1 min et une étape d'élongation 72°C pendant 20 secondes).

Ce fragment nucléotidique code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°8 : AIEKYLQDQA RLNSWGCAFR QVC.

- 20 • la reconnaissance des anticorps anti-HIV-1 (groupe M ; gp120) :

La séquence SEQ ID N°9 est issue de la souche virale HIV-1 groupe M (clone de référence HXB2) et correspond à la séquence suivante :

- SEQ ID N°9 : GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA
TTA TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA
25 CCC ACC AAG.

Cette séquence est amplifiée par PCR par l'utilisation d'amorces d'amplification spécifiques (amorce sense 5' GTC TGC TCG AGC GGT TCT GGA GGA GGA GAT ATG AGG 3' et amorce antisens 5' ACG TCC TGC AGA CTT GGT GGG TGC TAC TCC 3' ; 17 cycles de PCR sont réalisés comprenant à chaque cycle une étape de
30 dénaturation à 94°C pendant 1min, une étape d'hybridation à 52°C pendant 1 min et une étape d'élongation à 72°C pendant 20 secondes).

Ce fragment nucléotidique code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°10 : GGGDMRDNWR SELYKYKVVK IEPLGVAPTK

b) des régions de liaisons, entre chacune des régions épitopiques citées précédemment, permettant :

5

- au niveau nucléotidique l'introduction de six sites de coupure par des enzymes de restriction pouvant être utilisés pour modifier, retirer ou ajouter un domaine épitopique, et
- au niveau protéique, l'obtention de régions d'espacement, flexibles, assurant

10

Ainsi, la séquence nucléotidique SEQ ID N°11 : *G* AGC TCC GGT TCA GGC permet l'obtention d'un site de coupure par l'enzyme Sac I (indiqué en gras), le *G* indiqué en italique, étant la dernière base de la séquence nucléotidique codant le peptide permettant la reconnaissance des anticorps anti HIV-1, groupe M. Cette séquence code la région flexible correspondant au peptide de séquence SEQ ID N°12 : SSG SG.

15

La séquence nucléotidique SEQ ID N°13 : *C* TCG AGC GGT TCT permet l'obtention d'un site de coupure par l'enzyme Xho I (indiqué en gras), le *C* indiqué en italique, étant la dernière base de la séquence nucléotidique codant le peptide permettant la reconnaissance des anticorps anti HIV-2. Cette séquence code la région flexible correspondant au peptide de séquence SEQ ID N°14 : SSGS.

20

La séquence nucléotidique SEQ ID N°15 : TCT GCA GGC permet l'obtention d'un site de coupure par l'enzyme Pst I (indiqué en gras). Cette séquence code la région flexible correspondant au peptide de séquence SEQ ID N° 16 : SAG

25

La séquence nucléotidique SEQ ID N°17 : AAA GCT TCC permet l'obtention d'un site de coupure par l'enzyme Hind III (indiqué en gras). Cette séquence code la région flexible correspondant au peptide de séquence SEQ ID N° 18 : KAS.

30

Les séquences SEQ ID N°19 : ATG AGG GGA TCC et SEQ ID N°20 TGA TCT AGA permettent respectivement d'obtenir un site de coupure par l'enzyme BamHI (indiqué en gras) et un site de coupure par l'enzyme XbaI permettant l'insertion ou l'extraction de la totalité de la séquence codant la protéine recombinante selon l'invention dans un plasmide.

c) **une région de fixation et permettant la purification de la protéine recombinante chimérique** : une séquence hexahistidine est ajoutée en C-terminal afin de faciliter ultérieurement l'étape de purification de la protéine recombinante chimérique. Ce peptide, codée par la séquence nucléotidique SEQ ID N°21 : CAC CAT CAC CAT CAC CAT, correspond à la séquence SEQ ID N°22 : HHHHHH.

A titre indicatif, cette région de fixation particulière, comprenant une succession d'histidine, permet notamment la fixation orientée de la protéine recombinante sur un support constitué de silice ou d'oxydes métalliques, tel que décrit dans le brevet FR-B-98/04879.

L'ordre des séquences codant les différentes régions épitopiques immunodominantes de la protéine recombinante chimérique peut être éventuellement modifié. Certains épitopes peuvent être présentés plusieurs fois au sein de la protéine recombinante chimérique. Les épitopes peuvent également présenter des variations par rapport aux séquences décrites dans l'exemple ci dessus selon le sous type ou le clone HIV qu'ils représentent. La longueur des régions de liaison peut également être modifiée pour améliorer l'accessibilité d'un épitope. Enfin, les régions de fixation peuvent être insérées au sein des régions de liaisons.

Ainsi, les inventeurs ont mis en évidence que des séquences épitopiques plus courtes que celles décrites précédemment peuvent également être utilisées. Ces séquences permettent :

- la reconnaissance des anticorps anti-HIV-1 (groupe M), :

La séquence SEQ ID N°27 est issue de la souche virale HIV-1 groupe M (clone de référence HXB2) et correspond à la séquence suivante :

SEQ ID N°27 : GAA AGA TAC CTA AAG GAT CAA CAG CTC CTA GGG ATT
TGG GGT TGC TCT GGA AAA CTC ATT TGC ACC ACG

Ce fragment nucléotidique code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°28 : ERYLKDQQLL GIWGCSGKLI CTT.

- la reconnaissance des anticorps anti-HIV-1 (groupe O ; gp41) :

La séquence SEQ ID N° 29 correspond à une séquence ADN artificielle, conçue à partir de la séquence en acide aminé de la souche virale HIV-1 groupe O [clone ANT70]. Cette portion synthétique a été conçue en sélectionnant des codons dont l'utilisation est favorable à l'expression de gène chez *E. coli*. La séquence est la suivante :

5 SEQ ID N° 29 : GAA ACC CTG CTT CAG AAC CAA CAG CTG CTT TCT CTG
TGG GGT TGC AAA GGT AAG CTG GTT TGC TAC ACC

Ce fragment nucléotidique code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°30 : ETLLQNQQLL SLWGCKGKLV CYT.

• la reconnaissance des anticorps anti-HIV-2 (gp36) :

10 La séquence SEQ ID N°31 est issue de la souche virale HIV-2 (clone de référence ROD) et correspond à la séquence suivante :

SEQ ID N°31 : CTA AAT TCA TGG GGA TGT GCG TTT AGA CAA GTC TGC

Ce fragment nucléotidique code pour le peptide correspondant à la séquence en acides aminés SEQ ID N°32 : LNSWGCAFR QVC

15

De plus, les inventeurs ont également utilisés les régions de liaisons suivantes :

- la séquence nucléotidique SEQ ID N°41 CTG CAC CAT ATC CTG GAA GCC CAG CGT ATG GAA TGG CAC CCG CAC AAA GGT TCT GGA TCC qui correspond à la séquence en acides aminés SEQ ID N°42 LHHILEAQRM EWHPHKGS
- 20 • la séquence nucléotidique SEQ ID N°43 CTG CAC CAT ATC CTG GAG GCT CAA CGT ATG GAG TGG CGC GAA TCC CAT GGT qui correspond à la séquence en acides aminés SEQ ID N°44 LHHILEAQRM EWRESHG
- la séquence nucléotidique SEQ ID N° 45 GGT CTG AAC GAC ATC CTG
25 GAA GCC CAG CGT ATG GAA TGG CAC GAG TCT GCA GGC qui correspond à la séquence en acides aminés SEQ ID N°46 GLKDILEAQR MEWHESAG
- la séquence nucléotidique SEQ ID N°47 CTG AAC GAT ATT TTC GAA GCG CAG CGT ATT GAA TGG CAT GAG GGT TCT GGA TCC qui correspond à
30 la séquence en acides aminés SEQ ID N°48 LNDIFEAQRI EWHEGSGS

Le remplacement d'une arginine par une lysine au sein d'une région de liaison telle que définie ci dessus permet la fixation d'une biotine qui permet d'obtenir une région de fixation au sein de la région de liaison.

Les inventeurs ont ainsi utilisés les séquences suivantes qui permettaient non seulement
5 de lier les régions épitopiques entre elles mais également de fixer la protéine recombinante chimérique selon l'invention sur un support, ou de faciliter sa purification.

Ainsi, les inventeurs ont également utilisés les régions de liaisons suivantes :

- 10 • la séquence nucléotidique SEQ ID N°33 CTG CAC CAT ATC CTG GAA GCC CAG AAA ATG GAA TGG CAC CCG CAC AAA GGT TCT GGA TCC qui correspond à la séquence en acides aminés SEQ ID N°34 LHHILEAQKM EWHPHKSGS
- 15 • la séquence nucléotidique SEQ ID N°35 CTG CAC CAT ATC CTG GAG GCT CAA AAG ATG GAG TGG CGC GAA TCC CAT GGT TCC CAT GGT qui correspond à la séquence en acides amines SEQ ID N°36 LHHILEAQKM EWRESHG
- la séquence nucléotidique SEQ ID N°37 GGT CTG AAC GAC ATC CTG GAA GCC CAG AAA ATG GAA TGG CAC GAG TCT GCA GGC qui correspon à la SEQ ID N°38 GLKDILEAQK MEWHESAG
- 20 • la séquence nucléotidique SEQ ID N°39 CTG AAC GAT ATT TTC GAA GCG CAG AAG ATT GAA TGG CAT GAG GGT TCT GGA TCC qui correspond à la séquence en acides aminés SEQ ID N°40 LNDIFEAQKI EWHEGSGS

Les séquences présentées ci dessus ont permis la construction de la protéine
25 recombinante chimérique bHIV86.

Ainsi, la séquence nucléotidique SEQ ID N°49 a été conçue pour coder une protéine recombinante selon l'invention, et clonée dans un vecteur d'expression.

SEQ ID N°49: ATG AGG GGA TCT CTG CAC CAT ATC CTG GAA GCC CAG
AAA ATG GAA TGG CAC CCG CAC AAA GGT TCT GGA TCC GAA AGA TAC
30 CTA AAG GAT CAA CAG CTC CTA GGG ATT TGG GGT TGC TCT GGA AAA
CTC ATT TGC ACC ACG AGC TCC CTG CAC CAT ATC CTG GAG GCT CAA

AAG ATG GAG TGG CGC GAA TCC CAT GGT CTA AAT TCA TGG GGA TGT
 GCG TTT AGA CAA GTC TGC TCG AGC GGT CTG AAG GAC ATC CTG GAA
 GCC CAG AAA ATG GAA TGG CAC GAG TCT GCA GGC GAA ACC CTG CTT
 CAG AAC CAA CAG CTG CTT TCT CTG TGG GGT TGC AAA GGT AAG CTG
 5 GTT TGC TAC ACC AAA GCT TCC CAC CAT CAC CAT CAC CAT TGA TCT
 AGA

La protéine recombinante chimérique codée par la séquence SEQ ID N°50 est la suivante :

SEQ ID N°50 : MRGSLHHILE AQKMEWHPHK GSGSERYLKD QQLLGIWGCS
 10 GKLICTTSSL HHILEAQKME WRESHGLNSW GCAFRQVCSS GLKDILEAQK
 MEWHESAGET LLQNQQLLSL WGCKGKLV CY TKASHHHHHH

Les régions épitopiques sont indiquées en gras, la région de fixation en italique et les régions de liaisons en non gras non italique.

15

De même, les séquences présentées précédemment ont permis la construction de la protéine recombinante chimérique bHIV98 dans laquelle la séquence hexa histidine a été déplacé en N terminal, afin de faciliter sa purification.

Ainsi, la séquence nucléotidique SEQ ID N°51 a été conçue pour coder une protéine
 20 recombinante selon l'invention, et clonée dans un vecteur d'expression.

SEQ ID N°51: ATG AGG GGA TCT CAC CAT CAC CAT CAC CAT GGT CTG
 AAC GAT ATT TTC GAA GCG CAG AAG ATT GAA TGG CAT GAG GGT TCT
 GGA TCC GAA AGA TAC CTA AAG GAT CAA CAG CTC CTA GGG ATT TGG
 GGT TGC TCT GGA AAA CTC ATT TGC ACC ACG AGC TCC CTG CAC CAT
 25 ATC CTG GAG GCT CAA AAG ATG GAG TGG CGC GAA TCC CAT GGT CTA
 AAT TCA TGG GGA TGT GCG TTT AGA CAA GTC TGC TCG AGC GGT CTG
 AAG GAC ATC CTG GAA GCC CAG AAA ATG GAA TGG CAC GAG TCT GCA
 GGC GAA ACC CTG CTT CAG AAC CAA CAG CTG CTT TCT CTG TGG GGT
 TGC AAA GGT AAG CTG GTT TGC TAC ACC TGA A GCT T

30

La protéine recombinante chimérique codée par la séquence SEQ ID N°52 est la suivante :

SEQ ID N°52 : MRGSHHHHHH GLNDIFEAK IEWHEGSGSE RYLKDQQLLG
IWGCSGKLIC TTSSLHHILE AQKMEWRESH GLNSWGCAFR QVCSSGLKDI
5 LEAQKMEWHE SAGETLLQNQ QLLSLWGCKG KLV CYT.

Les régions épitopiques sont indiquées en gras, la région de fixation en italique et les régions de liaisons en non gras non italique.

**Exemple 2 : Expression et purification des protéines recombinantes chimériques
b-HIV 72, b-HIV86 et b-HIV98 de l'exemple 1**

La première étape consiste à insérer la séquence SEQ ID N°1 (exemple 1), dans un vecteur d'expression (pMR) puis à transformer une bactérie *E. coli* (souche BL21) avec la construction plasmidique obtenue selon un protocole classique de clonage connu de l'homme du métier. Les bactéries transformées sont sélectionnées grâce à leur
15 résistance à l'ampicilline portée par le vecteur pMR.

Un clone de bactérie recombinante est alors sélectionné pour ensemercer une préculture de 40 ml de milieu 2x YT (tryptone 16 g/l ; yeast extract 10 g/l ; NaCl 5 g/l, pH 7,0) contenant 100 µg/ml ampicilline. Après 15 à 18h d'incubation à 37°C sous agitation à 250 rpm, cette préculture est utilisée pour ensemercer 1 litre de milieu 2xYT contenant
20 2 % glucose et 100 µg/ml ampicilline. Cette culture est incubée à 37°C sous agitation à 250 rpm. Lorsque la $DO_{600\text{ nm}}$ a atteint 0,7-0,9, de l'IPTG (isopropyl-β-D-thiogalactoside, Eurogentec) est ajouté à 0,5 mM final dans le milieu de culture et la culture est poursuivie pendant 4 h. L'IPTG permet d'induire l'expression de la protéine chimérique recombinante SEQ ID N°2, N°50 ou N°52 qui s'accumule dans les
25 bactéries sous forme de corps d'inclusion. Après 4 h d'induction, la culture est centrifugée à 6000 rpm pendant 30 min à 4°C et le culot de bactéries est congelé à -80°C.

Afin d'extraire la protéine recombinante des corps d'inclusion, les bactéries décongelées sont lysées. Pour cela, les culots de bactéries correspondant à une culture
30 d'un litre sont repris dans 100 ml de tampon de lyse (PBS 1X contenant des inhibiteurs de protéases : lysozyme : 1 mg/ml ; benzonase : 2,5 unités par ml (Novagen®) et Mg^{2+} :

1 mM) en vortexant jusqu'à obtenir une suspension homogène. Cette solution est incubée 1 heure à température ambiante sous agitation. La solution est ensuite centrifugée 30 min à 4°C à 10000 g.

Le culot obtenu contient les corps d'inclusion. Ce culot est mis en suspension dans 50
5 ml de tampon de solubilisation (bicarbonate de sodium : 40 mM ; NaCl : 300 mM ; SDS : 1% ; β -mercaptoethanol : 20 mM, pH 9,6) contenant des inhibiteurs de protéases (complete EDTA-free, Roche®). La solution ainsi obtenue est incubée 16 à 18 h sous agitation entre 18 et 25°C. Elle est ensuite diluée au quart avec un tampon PBS 2X contenant 8 mM d'imidazole et des inhibiteurs de protéases (complete EDTA-free,
10 Roche®) à pH 8,0. Une centrifugation à 10 000 g pendant 30 min à 20°C permet d'obtenir un surnageant limpide, qui est filtré sur filtre 0,45 μ et purifié par chromatographie d'affinité sur une colonne de chélation de métaux (matrice nickel-nitrilotriacetic acid (Ni-NTA, Qiagen). Les 200 ml d'échantillon sont déposés (1 ml/min) à 18-25°C sur une colonne de 8 ml de gel Ni-NTA équilibrée en tampon A1
15 (PBS 2X, urée 4 M, imidazole 6 mM, pH 7,8 contenant β -mercaptoethanol 5 mM) ou A2 (PBS 2X, SDS 0.25 %, imidazole 6 mM, pH 7,8 contenant β -mercaptoethanol 5 mM). La colonne est ensuite lavée en tampon A1 ou A2, jusqu'à obtention en sortie de colonne d'une $DO_{280nm}=0$. L'élution de la protéine recombinante est obtenue par application d'un tampon B1 (PBS 2X, urée 4M, imidazole 100 mM, pH 7,5, contenant
20 β -mercaptoethanol 5 mM) ou B2 (PBS 2X, SDS 0.25%, imidazole 100 mM, pH 7,5, contenant β -mercaptoethanol 5 mM).

Des quantités de l'ordre de 50 mg de protéine recombinante purifiée peuvent être obtenues à partir d'un litre de culture.

La protéine recombinante ainsi purifiée est soumise à un traitement dénaturant par
25 addition de SDS (1500 molécules par molécule de protéine recombinante), DTT 5 mM, bicarbonate de sodium 50 mM au pH 9,6 et chauffage 30 min à 37°C. La stoechiométrie molécules de SDS/ molécules de protéine recombinante peut être modifiée (idéalement diminuée si la durée ou température de chauffage est augmentée). Par exemple, des résultats similaires sont obtenus par addition de 250 molécules de SDS/molécule de
30 protéine recombinante, DTT 5 mM, bicarbonate de sodium 50 mM au pH 9.6 et chauffage 2 heures à 40°C.

La protéine ainsi dénaturée est stabilisée par addition de Polyéthylène glycol (MW 3350 en particulier) pour une stockiométrie de 10 molécules de PEG par molécule de protéine puis dialysée à 4°C pendant 18 à 24 heures contre un tampon bicarbonate de sodium 50mM, EDTA 1mM, SDS 0,01%, PEG 1 mg/l, pH 9,6.

5 L'expression et la purification des protéines recombinantes chimériques b-HIV86 et b-HIV98 ont été réalisées d'une manière comparable, à l'exception de l'étape de dénaturation par SDS et DTT, qui n'est pas nécessaire lors de la purification des protéines b-HIV86 et b-HIV98. L'emploi de beta mercaptoéthanol dans les tampons de purification n'est pas nécessaire non plus.

10

Exemple 3 : Evaluation et validation de la protéine recombinante chimérique b-HIV 72, bHIV 86 et b-HIV98 dans un test VIDAS® (bioMérieux)

Cette validation est réalisée en test VIDAS® par l'utilisation d'une solution de protéine chimérique recombinante obtenue selon les exemples 1 et 2 et ayant subi le traitement
15 dénaturant décrit dans l'exemple 2.

Le principe du test VIDAS® est le suivant: un cône constitue le support solide qui sert également de système de pipetage pour les réactifs présents dans la barrette. La protéine recombinante est fixée sur le cône. Après une étape de dilution, l'échantillon est aspiré et refoulé plusieurs fois à l'intérieur du cône. Ceci permet aux IgG anti-VIH de
20 l'échantillon de se lier à la protéine recombinante. Les composants non liés sont éliminés par lavage. Un anticorps anti-IgG humaines conjugué à la phosphatase alcaline (PAL) est alors incubé dans le cône où il se fixe aux IgG anti-VIH. Des étapes de lavage éliminent le conjugué non fixé.

Pendant la dernière étape de révélation, le substrat de la PAL, le 4-méthyl-ombelliferyl phosphate, est hydrolysé en 4-méthyl-ombelliférone dont la fluorescence émise à 450 nm est mesurée. L'intensité de la fluorescence est mesurée par le système optique du Vidas® et est proportionnelle à la présence d'IgG anti-VIH présents dans l'échantillon. Les résultats sont analysés automatiquement par le VIDAS® et exprimés en RFV (Relative Fluorescent Value).

30 Dans cet exemple, une solution de protéine recombinante obtenue selon les exemples 1 et 2, (1,2 µg dans un millilitre de tampon bicarbonate de sodium 50 mM, SDS 0,01%,

pH 9,6-9,8) est incubée avec les cônes VIDAS® 18 à 24 h à température ambiante (120 µl/cône). Les cônes sont ensuite incubés dans un tampon de passivation (330 µl/cône de tampon HIV Duo contenant 3 % de sérum de veau) pendant 18 à 24 h à température ambiante

- 5 Des solutions tests (Etablissement Français du Sang, France), de sérologie HIV connue, (28 µl de sérum, de statut HIV connu, dilué dans 300 µl de tampon PBS 1 X, NaCl 8,76 g/l, tween 20 2,5 % (v/v), lait écrémé en poudre 2,5 g/l, albumine 20 g/l, et 3 % (v/v) sérum de veau, pH 6,1) sont alors mis en contact avec les cônes présentant les protéines recombinantes de l'exemple 1, pendant 13 min et 20 secondes (80 cycles de
- 10 pipettage/refoulement de 10 secondes). Une étape de lavage est ensuite effectuée en tampon Tris 24,23 g/l, acide maléique 23,22 g/l, tween 20 0,05% (v/v), NaOH 6 g/l, NaCl 8,77 g/l, pH 6,1. Une solution d'anticorps anti-Fc humain conjugué à la PAL (P5F2F7) et dilué au 1/5000^{ème} est incubée au contact du cône pendant 5 min (avec 30 cycles de pipettage /refoulement de 10 secondes chacun). Une dernière étape de lavage
- 15 est réalisé en tampon HIV Duo, avant l'étape finale de révélation.

Les résultats obtenus, sont exprimés en RFV (Relative Fluorescent Value). Les valeurs RFV supérieures ou égales à 250 sont arbitrairement considérées comme provenant d'un sérum HIV séro positifs. Les valeurs inférieures sont négatives. Les résultats obtenus avec la protéine recombinante obtenue selon les exemples 1 et 2, sont tous en

20 accord avec la sérologie HIV, préalablement déterminée par le test VIDAS HIV Duo® (bioMérieux®).

L'utilisation d'une telle protéine recombinante en test VIDAS® permet bien de déterminer la sérologie HIV des échantillons.

Tableau 1 – Validation expérimentale de la protéine recombinante b-HIV72 obtenue selon les exemples 1 et 2

Solution calibrateur / sera	Références sérum	RFV (relative fluorescent value)
Sera HIV1-M positifs	9991574	9820
	9991504	9346
	9991544	9621
	9991524	9735
	9991500	9914
Sera HIV1-O positifs	48	786
Sera HIV 2 positifs	7312A	9174
	JS8-1002 0008	8475
	JS8-1002 0009	9532
	JS8-1002 0010	9472
	JS8-1002 0013	9526
	JS8-1002 0014	9695
	JS8-1002 0016	9872
	JS8-1002 0017	9090
	JS8-1002 0018	9241
	JS8-1002 0020	9391
Sera HIV séronégatifs	203	139
	222	171
	262	105
	263	84
	280	182
	285	86
	287	127
	291	173
	g10653	193
	g12290	122
	g13461	170
	g13818	192
	g13830	106
	p03512	176

Des résultats comparables ont été obtenus avec les protéines recombinantes chimériques b-HIV86 et b-HIV98.

Exemple 4 : Construction d'une protéine recombinante biotinylée

5 Une séquence consensus de biotinylation *in vivo* dans *E.coli* telle que décrite par Schatz, (Bio/technology, vol. 11, 1993) peut être fusionnée à la SEQ ID n°2

L'addition de la séquence SEQ ID N°23 : 5' CTG CAC CAT ATC CTG GAA GCC CAG AAA ATG GAA TGG CAC CCG CAC, codant le peptide de séquence SEQ ID N°24 : LHHILEAQKM EWHPH, permet la biotinylation de la protéine recombinante b

10 HIV72 obtenue selon les exemples 1 et 2 afin d'utiliser une protéine avidine conjuguée à une enzyme pour la phase de révélation dans un test EIA.

Les conditions d'expression et purification décrites dans l'exemples 2 restent valides avec quelques modifications. Les bactéries transformées avec le plasmide recombinant

15 peuvent être de type BL21 ou AVB101 (Avidity, LLC). Le milieu de culture utilisé pour l'expression peut être de type : 2x YT (tryptone 16 g/l ; yeast extract 10 g/l ; NaCl 5 g/l, pH 7,0) contenant 100 µg/ml ampicilline et supplémenté par 12 µg par ml de biotine.

20 **Exemple 5 : Evaluation et validation de la protéine recombinante chimérique biotinylée *in vivo* tel que décrite dans l'exemple 4 dans un test VIDAS® de type sandwich.**

Cette validation est selon un protocole VIDAS® décrit dans l'exemple 3 et modifié comme suit : la protéine chimérique recombinante bHIV-72, non biotinylée obtenue

25 selon l'exemple 1 est fixée sur la phase solide (cône VIDAS®) comme décrit dans l'exemple 3. Après incubation de l'échantillon dilué avec le cône puis lavage, les Ig G anti HIV liées aux cônes sont incubées avec une protéine chimérique recombinante biotinylée *in vivo* et obtenue selon l'exemple 4. Après lavage, les protéines recombinantes biotinylées fixées sur le cône réagissent avec une solution de

30 streptavidine conjugué à la PAL. L'étape finale de révélation est conforme à la description de l'exemple 3.

Les résultats sont présentés dans le tableau 2. Les valeurs RFV supérieures ou égales à 250 sont arbitrairement considérées comme positives. L'utilisation d'une telle protéine recombinante en test VIDAS® permet bien de déterminer la sérologie HIV des échantillons.

5

Tableau 2 – Validation expérimentale de la protéine recombinante obtenue selon les exemples 4 et 5

Solution calibrateur / sera	Références sérum	RFV
Séra HIV négatifs	CTS 8	195
	g06976	214
	g31377	170
Séra HIV 1 M positifs	9991574	4708
	9991504	1395
Séra HIV 2 positif	JS8-1002 0008	2562
	JS8-1002 0009	3754

10 **Exemple 6 : Evaluation et validation des protéine recombinante chimérique bHIV-86 et bHIV-98 dans un test VIDAS® de type sandwich.**

La présence de la séquence SEQ ID N°33, 35 37 ou 39 codant respectivement les peptides SEQ ID N°34, 36, 38 et 40 permet aux protéines bHIV86 et bHIV98 d'être biotinylées in vivo, d'une façon comparable a ce qui est décrit dans l'exemple 4.

15 Cette validation est selon un protocole VIDAS® décrit dans l'exemple 3 et modifié comme suit : la protéine chimérique recombinante bHIV-72, non biotinylée obtenue selon l'exemple 1 est fixée sur la phase solide (cône VIDAS®) comme décrit dans l'exemple 3. Après incubation de l'échantillon dilué avec le cône puis lavage, les Ig G
20 anti HIV liées aux cônes sont incubées avec une protéine chimérique recombinante biotinylée in vivo (bHIV-86 ou bHIV-98). Après lavage, les protéines recombinantes biotinylées fixées sur le cône réagissent avec une solution de streptavidine conjugué à la PAL. L'étape finale de révélation est conforme à la description de l'exemple 3.

Des résultats comparables à ceux décrits dans l'exemple 5 ont été obtenues avec les protéines bHIV-86 et bHIV-98.

Exemple 7 : Amélioration de la sensibilité de la protéine recombinante chimérique

- 5 Afin d'augmenter encore la sensibilité de reconnaissance des anticorps anti-HIV de la protéine recombinante obtenue selon l'invention de nouveaux épitopes caractéristiques de certains sous-types HIV peuvent être ajoutés ou un des épitopes décrit peut être dupliqué dans la séquence de la protéine chimérique.

- 10 **Exemple 8 : Addition d'une séquence hexalysine à la protéine recombinante pour faciliter le couplage d'une enzyme ou de biotine, notamment sur la fonction β -aminé de la lysine.**

- Une séquence codant pour six lysines peut être fusionnée en 3' de la SEQ ID n°2. L'addition de la séquence ADN SEQ ID n°25: AAG AAA AAG AAA AAG AAA, codant pour le peptide de séquence SEQ ID n°26 : KKKKKK, permet le couplage
15 orienté d'enzyme ou biotine en C-terminal de la protéine chimérique. Le couplage de cette protéine recombinante à la phosphatase alcaline en particulier permet l'utilisation de la protéine couplée dans un format sandwich de révélation des anticorps anti-HIV.

REVENDICATIONS

1. ADN recombinant codant une protéine recombinante chimérique comprenant

- au moins deux premiers fragments nucléotidiques codant chacun une région épitopique du virus HIV-1 de groupe M ou de groupe O ou le virus HIV-2
- au moins un deuxième fragment nucléotidique codant une région de liaison,
- au moins un troisième fragment nucléotidique codant une région de fixation

caractérisé en ce que chaque premiers fragments nucléotidiques code au moins une région immunodominante de la glycoprotéine gp 120 d'HIV-1, de la glycoprotéine gp 41 d'HIV-1 de groupe M, de la glycoprotéine gp 41 d'HIV-1 de groupe O ou de la glycoprotéine gp 36 d'HIV-2.

2. ADN, selon la revendication 1, caractérisé en ce que ledit premier fragment nucléotidique a pour séquence l'une quelconque des séquences SEQ ID N°3, SEQ ID N°5, SEQ ID N°7, SEQ ID N°9, SEQ ID N°27, SEQ ID N°29 ou SEQ ID N°31.

3. ADN recombinant, selon l'une quelconque des revendications 1 ou 2, caractérisé en ce que ledit deuxième fragment nucléotidique comprend au moins un site de coupure.

4. ADN recombinant, selon l'une quelconque des revendications 1 à 3, caractérisé en ce que ledit deuxième fragment nucléotidique a pour séquence au moins l'une quelconque des séquences suivantes, prises seules ou en combinaison, SEQ ID N°11 SEQ ID N°13, SEQ ID N°15, SEQ ID N°17, SEQ ID N°19, ou SEQ ID N°20, SEQ ID N°33, SEQ ID N°35, SEQ ID N°37, SEQ ID N°39, SEQ ID N°41, SEQ ID N°43 SEQ ID N°45, ou SEQ ID N°47

5 ADN recombinant selon l'une quelconque des revendications 1 à 4 caractérisé en ce que ledit troisième fragment nucléotidique codant une région de fixation est inclus dans ledit deuxième fragment nucléotidique codant une région de liaison.

5 6 ADN, selon l'une quelconque des revendications 1 à 5, caractérisé en ce que ledit troisième fragment nucléotidique a pour séquence l'une quelconque des séquences SEQ ID N°21, SEQ ID N° 23, SEQ ID N°25, SEQ ID N°33, SEQ ID N°35, SEQ ID N°37 ou SEQ ID N°39.

10 7. Protéine recombinante chimérique codée par un ADN recombinant, selon l'une quelconque des revendications 1 à 6, comprenant

- au moins deux régions épitopiques du virus HIV-1 de groupe M ou de groupe O ou le virus HIV-2,
- au moins une région de liaison
- 15 • au moins une région de fixation.

8. Protéine, selon la revendication 7, caractérisée en ce que ladite région de liaison est un peptide comprenant au moins une glycine et/ou au moins une sérine.

20 9. Protéine, selon la revendication 8, caractérisée en ce que ladite région de liaison a pour séquence l'une quelconque des séquences SEQ ID N° 12, SEQ ID N°14, SEQ ID N°16 ou SEQ ID N°18, 34, 36, 38, 40, 42, 44, 46, 48.

25 10. Protéine, selon la revendication 7, caractérisée en ce que ladite région de fixation est une région riche en histidines et ses dérivés, tel qu'une région contenant une densité d'histidines, supérieure ou égale à 25%, et préférentiellement supérieure ou égale à 33%.

30 11. Protéine selon la revendication 7 caractérisée en ce que ladite région de fixation est un peptide comprenant au moins une lysine.

12. Protéine selon la revendication 7 caractérisée en ce que ladite région de fixation a pour séquence SEQ ID N°22, 24, 26, 34, 36, 38 ou 40.

5

13. Vecteur d'expression comprenant un ADN recombinant selon l'une quelconque des revendications 1 à 6.

10

14 Utilisation d'au moins un ADN selon l'une quelconque des revendications 1 à 6 et/ou d'au moins une protéine recombinante chimérique selon l'une quelconque des revendications 7 à 12 pour le diagnostic in vitro.

SEQUENCE LISTING

<110> BIOMERIEUX SA

<120> Protéine recombinante chimérique et diagnostic in vitro

<130> Unknown

<160> 52

<170> PatentIn version 3.1

<210> 1

<211> 423

<212> DNA

<213> Artificial

<400> 1

```

atgaggggat ccagaatcct agctgtggaa agatacctaa aggatcaaca gctcctaggg      60
atttgggggtt gctctggaaa actcatttgc accactgctg tgagctccgg ttcaggcgct      120
atagagaagt acctacagga ccaggcgcgg ctaaattcat ggggatgtgc gtttagacaa      180
gtctgtctga gcggttctgg aggaggagat atgagggaca attggagaag tgaattatat      240
aaatataaag tagtaaaaat tgaaccatta ggagtagcac ccaccaagtc tgcaggccgt      300
ctgcttgctc tggaaaccct gcttcagaac caacagctgc tttctctgtg gggttgcaaa      360
ggtaagctgg tttgctacac ctctgttaaa gcttcccacc atcaccatca ccattgatct      420
aga                                                                    423

```

<210> 2

<211> 138

<212> PRT

<213> Artificial

<400> 2

```

Met Arg Gly Ser Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln
1           5           10           15

```

```

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
20           25           30

```

Ala Val Ser Ser Gly Ser Gly Ala Ile Glu Lys Tyr Leu Gln Asp Gln
 35 40 45

Ala Arg Leu Asn Ser Trp Gly Cys Ala Phe Arg Gln Val Cys Ser Ser
 50 55 60

Gly Ser Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr
 65 70 75 80

Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys
 85 90 95

Ser Ala Gly Arg Leu Leu Ala Leu Glu Thr Leu Leu Gln Asn Gln Gln
 100 105 110

Leu Leu Ser Leu Trp Gly Cys Lys Gly Lys Leu Val Cys Tyr Thr Ser
 115 120 125

Val Lys Ala Ser His His His His His His
 130 135

<210> 3

<211> 90

<212> DNA

<213> HIV 1

<400> 3

agaatcctag ctgtggaaag atacctaaag gatcaacagc tcctagggat ttgggggttg 60

tctggaaaac tcatttgac cactgctgtg 90

<210> 4

<211> 30

<212> PRT

<213> HIV 1

<400> 4

Arg Ile Leu Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly
 1 5 10 15

Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val
20 25 30

<210> 5

<211> 90

<212> DNA

<213> HIV 1

<400> 5
cgtctgcttg ctctggaaac cctgcttcag aaccaacagc tgctttctct gtgggggtgc 60
aaaggttaagc tggtttgcta caoctctgtt 90

<210> 6

<211> 30

<212> PRT

<213> HIV 1

<400> 6

Arg Leu Leu Ala Leu Glu Thr Leu Leu Gln Asn Gln Gln Leu Leu Ser
1 5 10 15

Leu Trp Gly Cys Lys Gly Lys Leu Val Cys Tyr Thr Ser Val
20 25 30

<210> 7

<211> 69

<212> DNA

<213> HIV 2

<400> 7
gctatagaga agtacctaca ggaccaggcg cggctaaatt catggggatg tgcgtttaga 60
caagtctgc 69

<210> 8

<211> 23

<212> PRT

<213> HIV 2

<400> 8

Ala Ile Glu Lys Tyr Leu Gln Asp Gln Ala Arg Leu Asn Ser Trp Gly
1 5 10 15

Cys Ala Phe Arg Gln Val Cys
20

<210> 9

<211> 90

<212> DNA

<213> HIV 1

<400> 9

ggaggaggag atatgagggg caattggaga agtgaattat ataaatataa agtagtaaaa 60

attgaaccat taggagtagc acccaccaag 90

<210> 10

<211> 30

<212> PRT

<213> HIV 1

<400> 10

Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr
1 5 10 15

Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys
20 25 30

<210> 11

<211> 16

<212> DNA

<213> Artificial

<400> 11

gagctccggt tcaggc

16

<210> 12

<211> 5

<212> PRT

<213> Artificial

<400> 12

Ser Ser Gly Ser Gly
1 5

<210> 13

<211> 13

<212> DNA

<213> Artificial

<400> 13

ctcgagcggt tct

13

<210> 14

<211> 4

<212> PRT

<213> Artificial

<400> 14

Ser Ser Gly Ser
1

<210> 15

<211> 9

<212> DNA

<213> Artificial

<400> 15

tctgcaggc

9

<210> 16

<211> 3

<212> PRT

<213> Artificial

<400> 16

Ser Ala Gly

1

<210> 17

<211> 9

<212> DNA

<213> Artificial

<400> 17

aaagcttcc

9

<210> 18

<211> 3

<212> PRT

<213> Artificial

<400> 18

Lys Ala Ser

1

<210> 19

<211> 12

<212> DNA

<213> Artificial

<400> 19

atgaggggat cc

12

<210> 20

<211> 17

<212> DNA

<213> Artificial

<400> 20

artcaatgag gggatcc

17

<210> 21

<211> 18

<212> DNA

<213> Artificial

<400> 21

caccatcacc atcaccat

18

<210> 22

<211> 6

<212> PRT

<213> Artificial

<400> 22

His His His His His His
1 5

<210> 23

<211> 45

<212> DNA

<213> Artificial

<400> 23

ctgcaccata tcttgaagc ccagaaaatg gaatggcacc cgcac

45

<210> 24

<211> 15

<212> PRT

<213> Artificial

<400> 24

Leu His His Ile Leu Glu Ala Gln Lys Met Glu Trp His Pro His
1 5 10 15

<210> 25

<211> 18

<212> DNA

<213> Artificial

<400> 25

aagaaaaaga aaaagaaa

18

<210> 26

<211> 6

<212> PRT

<213> Artificial

<400> 26

Lys Lys Lys Lys Lys Lys
1 5

<210> 27

<211> 69

<212> DNA

<213> HIV 1

<400> 27

gaaagatacc taaaggatca acagctccta gggatttggg gttgctctgg aaaactcatt 60

tgcaccacg 69

<210> 28

<211> 23

<212> PRT

<213> HIV 1

<400> 28

Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser
1 5 10 15Gly Lys Leu Ile Cys Thr Thr
20

<210> 29

<211> 69

<212> DNA

<213> HIV 1

<400> 29

gaaaccctgc ttcagaacca acagctgctt tctctgtggg gttgcaaagg taagctgggt 60

tgctacacc 69

<210> 30

<211> 23

<212> PRT

<213> HIV 1

<400> 30

Glu Thr Leu Leu Gln Asn Gln Gln Leu Leu Ser Leu Trp Gly Cys Lys
1 5 10 15

Gly Lys Leu Val Cys Tyr Thr
20

<210> 31

<211> 36

<212> DNA

<213> HIV 2

<400> 31

ctaaattcat ggggatgtgc gttagacaa gtctgc

36

<210> 32

<211> 12

<212> PRT

<213> HIV 2

<400> 32

Leu Asn Ser Trp Gly Cys Ala Phe Arg Gln Val Cys
1 5 10

<210> 33

<211> 60

<212> DNA

<213> Artificial

<400> 33

ctgcaccata tcttggaagc ccagaaaatg gaatggcacc cgcacaaagg ttctggatcc

60

<210> 34

<211> 20

<212> PRT

<213> Artificial

<400> 34

Leu His His Ile Leu Glu Ala Gln Lys Met Glu Trp His Pro His Lys
1 5 10 15

Gly Ser Gly Ser
20

<210> 35

<211> 51

<212> DNA

<213> Artificial

<400> 35

ctgcaccata tcctggaggg tcaaaagatg gagtggcgcg aatcccatgg t 51

<210> 36

<211> 17

<212> PRT

<213> Artificial

<400> 36

Leu His His Ile Leu Glu Ala Gln Lys Met Glu Trp Arg Glu Ser His
1 5 10 15

Gly

<210> 37

<211> 54

<212> DNA

<213> Artificial

<400> 37

ggctctgaagg acatcctgga agcccagaaa atggaatggc acgagtctgc aggc

54

<210> 38

<211> 18

<212> PRT

<213> Artificial

<400> 38

Gly	Leu	Lys	Asp	Ile	Leu	Glu	Ala	Gln	Lys	Met	Glu	Trp	His	Glu	Ser
1				5					10					15	

Ala Gly

<210> 39

<211> 54

<212> DNA

<213> Artificial

<400> 39

ctgaacgata ttttcgaagc gcagaagatt gaatggcatg agggttcttg atcc

54

<210> 40

<211> 18

<212> PRT

<213> Artificial

<400> 40

Leu	Asn	Asp	Ile	Phe	Glu	Ala	Gln	Lys	Ile	Glu	Trp	His	Glu	Gly	Ser
1				5					10					15	

Gly Ser

<210> 41

<211> 60

<212> DNA

<213> Artificial

<400> 41

ctgcaccata tcttggagc ccagcgtatg gaatggcacc cgcacaaagg ttctggatcc 60

<210> 42

<211> 22

<212> PRT

<213> Artificial

<400> 42

Leu His His Ile Leu Glu Ala Gln Arg Met Glu Trp His Pro His Pro
1 5 10 15

His Lys Gly Ser Gly Ser
20

<210> 43

<211> 51

<212> DNA

<213> Artificial

<400> 43

ctgcaccata tcttggagc tcaacgtatg gattggcgcg aatcccatgg t 51

<210> 44

<211> 17

<212> PRT

<213> Artificial

<400> 44

Leu His His Ile Leu Glu Ala Gln Arg Met Glu Trp Arg Glu Ser His
1 5 10 15

Gly

<210> 45

<211> 54

<212> DNA

<213> Artificial

<400> 45

ggtctgaagg acatcctgga agcccagcgt atggaatggc acgagtctgc aggc 54

<210> 46

<211> 18

<212> PRT

<213> Artificial

<400> 46

Gly Leu Lys Asp Ile Leu Glu Ala Gln Arg Met Glu Trp His Glu Ser
1 5 10 15

Ala Gly

<210> 47

<211> 54

<212> DNA

<213> Artificial

<400> 47

ctgaacgata ttttcgaagc gcagcgtatt gaatggcatg agggttctgg atcc 54

<210> 48

<211> 18

<212> PRT

<213> Artificial

<400> 48

Leu Asn Asp Ile Phe Glu Ala Gln Arg Ile Glu Trp His Glu Gly Ser
1 5 10 15

Gly Ser

<210> 49

<211> 399

<212> DNA

<213> Artificial

<400> 49

atgaggggat ctctgcacca taccctggaa gccagaaaa tggaatggca cccgcacaaa 60
ggttctggat ccgaaagata cctaaaggat caacagctcc tagggatttg gggttgctct 120
ggaaaactca ttgcaccac gagctccctg caccatatcc tggaggctca aaagatggag 180
tggcgcgaat cccatggtct aaattcatgg ggatgtgcgt ttagacaagt ctgctcgagc 240
ggcttgaagg acatcctgga agcccagaaa atggaatggc acgagtctgc aggcgaaacc 300
ctgcttcaga accaacagct gctttctctg tggggttgca aaggtaagct ggtttgctac 360
accaaagctt cccaccatca ccatcaccat tgatctaga 399

<210> 50

<211> 130

<212> PRT

<213> Artificial

<400> 50

Met Arg Gly Ser Leu His His Ile Leu Glu Ala Gln Lys Met Glu Trp
1 5 10 15

His Pro His Lys Gly Ser Gly Ser Glu Arg Tyr Leu Lys Asp Gln Gln
20 25 30

Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ser
35 40 45

Ser Leu His His Ile Leu Glu Ala Gln Lys Met Glu Trp Arg Glu Ser
50 55 60

His Gly Leu Asn Ser Trp Gly Cys Ala Phe Arg Gln Val Cys Ser Ser
65 70 75 80

Gly Leu Lys Asp Ile Leu Glu Ala Gln Lys Met Glu Trp His Glu Ser
85 90 95

Ala Gly Glu Thr Leu Leu Gln Asn Gln Gln Leu Leu Ser Leu Trp Gly
100 105 110

Cys Lys Gly Lys Leu Val Cys Tyr Thr Lys Ala Ser His His His His
115 120 125

His His
130

<210> 51

<211> 386

<212> DNA

<213> Artificial

<400> 51

atgaggggat ctcacccatca ccatcaccat ggtctgaacg atatttttcga agcgcagaag 60

attgaatggc atgaggggttc tggatccgaa agatacctaa aggatcaaca gctcctaggg 120

atttgggggtt gctctggaaa actcatttgc accacgagct ccctgcacca tatcctggag 180

gctcaaaaaga tggagtggcg cgaatcccat ggtctaaatt catgggggatg tgcgtttaga 240

caagtctgct cgagcgggtct gaaggacatc ctggaagccc agaaaatgga atggcacgag 300

tctgcaggcg aaaccctgct tcagaaccaa cagctgcttt ctctgtgggg ttgcaaaggt 360
 aagctgggttt gctacacctg aagctt 386

<210> 52

<211> 126

<212> PRT

<213> Artificial

<400> 52

Met Arg Gly Ser His His His His His His Gly Leu Asn Asp Ile Phe
 1 5 10 15

Glu Ala Gln Lys Ile Glu Trp His Glu Gly Ser Gly Ser Glu Arg Tyr
 20 25 30

Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu
 35 40 45

Ile Cys Thr Thr Ser Ser Leu His His Ile Leu Glu Ala Gln Lys Met
 50 55 60

Glu Trp Arg Glu Ser His Gly Leu Asn Ser Trp Gly Cys Ala Phe Arg
 65 70 75 80

Gln Val Cys Ser Ser Gly Leu Lys Asp Ile Leu Glu Ala Gln Lys Met
 85 90 95

Glu Trp His Glu Ser Ala Gly Glu Thr Leu Leu Gln Asn Gln Gln Leu
 100 105 110

Leu Ser Leu Trp Gly Cys Lys Gly Lys Leu Val Cys Tyr Thr
 115 120 125